

一次训练耗电2400万度

AI发展正在消耗大量能源!

今天我们常说的人工智能(AI),主要指的都是生成式人工智能。其中一大部分,是基于大语言模型的生成式AI。而让各种AI运转起来的服务器就是电子计算机,虽然计算机中处理的只是信息,但运转却需要大量能源。比如,仅仅是ChatGPT-4的GPU(计算机显卡)的核心部件,负责执行复杂的数学和几何计算,用于渲染图像、视频和其他图形内容,也用于AI的训练和推理,训练一次就会用去2400万度电。

AI带来的能源问题已经显现。



电子产品正消耗大量能源 制图 小翰

一个被证实的原理——处理信息需要消耗能量

1961年,在IBM公司工作的物理学家兰道尔·兰道尔提出了后来被称为“兰道尔原理”的理论。这一理论认为,计算机中存储的信息在发生不可逆的变化时,会向周围环境散发一点热量,其散发的热量和计算机当时所处的温度有关——温度越高,散发的热量越多。

兰道尔原理连接起了“信息”和“能量”,简单地说,就是信息处理操作,会导致物理世界中熵(热力学中表征物质状态的参量之一)的增加,从而消耗能量。

这一原理自提出以来曾遭受过不少质疑。但是近年来,兰道尔原理已被实验证明。2012年,《自然》杂志上发表了一篇文章,研究团队首次测量到了“位”(bit,是计算机中最基本的信息单位,它可以表示二进制的0或1)数据被删除时释放的微量热量。后来又有几次独立实验,都证明了兰道尔原理。所以,处理信息是有能量成本的。现在的电子计算机在计算时实际消耗的能量,甚至是这个理论值的数百倍。

大语言模型需要大量计算。它的工作过程大致可以分为训练和推理两个阶段。无论是训练还是推理,都是一连串信息重组的过程,因此都遵循兰道尔原理。于是不难推知,模型的参数量越大,需要处理的数据越多,所需的计算量就越大,所消耗的能量也就越大,释放的热量也就越多。

AI的能耗更多还是来自电流

科学家们一直在努力寻找更高效的计算方法,以降低成本。不过从目前的研究进展情况来看,也许只有真正的室温超导材料能广泛应用于计算设备时,才有可能降低这种消耗。

而这还只是AI耗电中微不足道的一小部分。更大的消耗则来自电流。

今天的电子计算机建立在集成电路的基础上。我们通常把集成电路叫做芯片,每个芯片中,又有许多晶体管。不严格地描述,晶体管可以理解成微小的开关。这些开关串联或者并联在一起,就可以实现逻辑运算。

“开”和“关”是表示两种状态,也就是所谓的1和0,就是计算的基本单位“位”。它是计算机二进制的基础。而计算机通过快速改变电压,来

拨动这些开关;改变电压,需要电子流入或流出;电子的流入流出,就构成了电流;又因为在电路中总是有电阻,就产生了热量。这个过程中产生的热量与电流的平方成正比,与导体电阻成正比,与通电时间成正比。

集成电路技术发展到现在,芯片中的晶体管已经变得极小。所以,单个晶体管所产生的热量并不会太高。但问题是,如今芯片上的晶体管实在是已经多到了常人无法想象的程度——比如,在IBM前几年发布的等效2纳米制程芯片中,每平方毫米面积上,平均有3.3亿个晶体管。

再小的热量,乘上这个规模,结果也一定相当可观。一个可能让人大跌眼镜的事实是,今天芯片单位体积的功率,比太阳核心的功

率还多出好几个数量级——典型的CPU芯片功率大概是每立方厘米100瓦,即每立方米1亿瓦;而太阳核心的功率只有每立方米不到300瓦。

在OpenAI公司训练大语言模型ChatGPT-4时,完成一次训练需要约三个月时间,要使用大约25000块“英伟达A100 GPU”,每块GPU都拥有540亿个晶体管,功耗400瓦,每秒钟可以进行19.5亿次单精度浮点数的运算,每次运算又涉及许多个晶体管的开关。

仅仅是这些GPU,一次训练就要用掉2400万度电。这些电能最后几乎全部转化成了热能,而这些能量可以将大约20万立方厘米冰水——大概是100个奥运会标准游泳池的水量——加热到沸腾。

AI发展带来的能源短缺问题已开始浮现

为什么AI需要用这么多的强大GPU来训练?因为大语言模型的规模实在太大了。

ChatGPT-3模型已拥有1750亿参数,而据推测,ChatGPT-4拥有1.8万亿参数,是ChatGPT-3的十倍。要训练这种规模的模型,需要大量计算。这些计算最终会表现为晶体管的开开关闭,和集成电路中细细的电流——以及热量。

而能量无法创造也无法消灭,它只能从一种形式转化成另一种形式。对于电子计算机来说,它主要是在将电能转化成热能。大语言模型也是如此,它对电能和冷却水的需求,正带来越来越严重的环境问题。

早前,有微软公司的工程师曾透露,为了训练ChatGPT-6,微软和OpenAI建造了巨大的数据中心,将会使用10万块“英伟达H100

GPU”——其性能比“英伟达A100 GPU”更强,当然功耗也更大。其中一个关键信息是,这些GPU不能放在同一个州,否则会导致电网负荷过大而崩溃。

在今年的达沃斯世界经济论坛上,OpenAI的CEO山姆·阿尔特曼认为,核聚变可能是能源的发展方向。但要开发出真正可用的核聚变技术,可能还需要一些时间。

水资源的问题也是一样。过去几年,那些在AI大模型领域先行一步的大企业,都面临水消耗大幅增长的局面。2023年6月,微软公司发布了2022年度环境可持续发展报告,其中用水一项,有超过20%的显著增长。谷歌公司也类似。

有研究者认为,AI的发展,是这些科技巨头用水量剧增的主要原因——要冷却疯狂发热的芯片,水冷系统是最

常见的选择。所以我们说,为AI提供硬件基础的数据中心,如同一个巨大的“电热水器”。

所以,AI发展带来的能源短缺问题,已经开始浮现。

如何让这些散失的热量不至于白白浪费?最容易想到也容易实现的,是热回收技术。比方说,将数据中心回收的热量用于提供民用热水,冬季提供民用采暖。现在有一些企业已经着手在回收废热再利用了,例如中国移动哈尔滨数据中心、阿里巴巴千岛湖数据中心等。

但这并不能从根本上解决问题。AI产业的发展速度之快,在人类历史上没有任何产业能与之相比。平衡AI技术的发展与环境的可持续性,可能会是我们未来几年的重要议题之一。技术进步和能源消耗之间的复杂关系,从来没有这么迫切地出现在人类面前。

本文由科普中国-星空计划出品,来源于微信公众号“返朴”(ID:fanpu2019)

每天吃5-6克洋葱,肝癌风险或降低33%

根据国际癌症研究机构(IARC)2020年全球癌症负担数据,2020年全球新发癌症病例1929万例,中国新发癌症病例457万例。对于肝癌来说,全球2020年新发肝癌91万例,其中中国新发肝癌41万例,占全球45%还要多。

世界癌症研究基金会(WCRF)第三次专家报告显示,食用受黄曲霉毒素污染的食物、喝酒会增加患肝癌的风险,而喝咖啡则可降低患肝癌的风险。

近日,上海交通大学医学院附属仁济医院项永兵等人在《欧洲流行病学杂志》上发表了一篇研究论文,称较高的洋葱和锰摄入量与肝癌风险降低相关——每天吃5-6克洋葱,肝癌风险或降低33%。

此外,较高的视黄醇摄入量则与肝癌风险增加相关;较高的脂肪、饱和脂肪、维生素B12摄入量与肝癌风险增加相关;而较高大米、韭菜、葱、叶类蔬菜摄入量则与肝癌风险降低相关。

研究人员表示,蔬菜富含膳食纤维、抗氧化剂和生物活性化合物,膳食纤维可减少肝脏脂肪的合成,减轻肝脏的代谢负担,还可以保持健康的肠道微生物系统,减少炎症和肝损伤的发生。

对于洋葱降低肝癌风险的机制,研究指出,洋葱与肝癌风险的负相关可归因于黄酮醇、二烯丙基硫醚和有机硫化物等成分。这些成分可能通过抗氧化特性、调节I和II相酶、刺激免疫系统以及抑制突变等,因而具有抑癌作用。

但该项研究还需进行多重比较和重复分析,未来,需要更多的流行病学研究来进一步验证。(浩源 整理)



资料图片

为流量,为发展

商界大佬“被迫营业”搞直播

刘强东、李彦宏、雷军等企业家,最近都瞄上了直播

堪称史上最密集“CEO直播”周

过去的一周,被网友称为史上最密集的“CEO直播”周。

4月16日,小米的雷军在微博宣布将在4月18日开启直播,“聊聊SU7开售这些天”。

实际上,SU7上市以来,流量和话题度居高不下。作为最强“跨界营销人”雷军在社交平台不遗余力地为新车宣传。截至4月16日,雷军已在微博发布十八集“小米SU7答网友问”,走在商业宣传的第一线。不少网友喊话,让雷军开直播。

无独有偶,同在4月16日,京东创始人、董事局主席刘强东,则以“采销东哥”数字人的形象在京东APP开启直播首秀,饮下了国内互联网企业大佬数字人直播的“头啖汤”。

“采销东哥”亮相京东家电商、京东超市采销直播间,带货品类包括粮油、农产品、宠物用品等,最便宜的仅1分钱,直

播27分钟,观看人数破千万。

要知道在刘强东之前,已有众多企业家纷纷涉足直播带货领域,包括网易CEO丁磊、搜狐CEO张朝阳、携程创始人梁建章、格力董事长董明珠等。但用“逼真”接近本尊的数字人,似乎对一众网友来说还是新鲜事。

实际上,雷军、刘强东的直播,并非一时兴起,而是最近7天内企业家“直播潮”里的浪花点点。

4月15日,长城汽车董事长魏建军直播亲测无图全场景NOA智驾,在河北保定完成了166公里智驾测试;又是在同一天,百度创始人、董事长兼首席执行官李彦宏在与极越CEO夏一平则共同开启了一场“双CEO”直播,一面试驾极越智能驾驶,一面介绍百度Apollo纯视觉智驾、文心一言等AI技术。

李彦宏更在直播中直言,汽车机器人就是未来。

“以后的汽车机器人会学习你各种各样的信息、学习你的喜好,越来越懂你。”他还在直播中谈到了大模型上车等话题,并表示:“你真的可以把它(车)当一个人来看待。”

4月14日,奇瑞控股集团党委书记尹同跃逻辑思维的罗振宇一起直播,挑战长途高速高阶智驾,为即将发售的产品预热。

如果把时间轴拉到一个月前,CEO直播的参与者还有更多:3月14日,蔚来汽车CEO李斌在其个人账号上献上直播首秀;3月26日,360创始人周鸿祎与哪吒汽车CEO张勇同框亮相直播;3月28日,吉利控股集团董事长李书福携手东方甄选董事长俞敏洪,在吉利超级工厂的一辆银河E8上进行了直播……



4月15日,李彦宏在直播中称,汽车机器人就是未来 图/直播截图

羊城晚报记者 王丹阳

面子和里子哪个最重要?

除了超级主播外,近来直播赛道的又一新景象是一众车企与互联网企业居然在一周内开始“卷老板”,让企业大佬放下面子,下场跑营销了。

首创“大佬数字人”直播是“一鱼三吃”吗?

随着车企大佬相继涌入直播间,雷军的微博评论区也被网友的热情淹没,他们纷纷喊话雷军开直播。

实际上,2020年8月,小米10周年之际,雷军就曾开启直播带货首秀。在这场直播中,雷军直播带货首秀销售金额为2.1亿元,累计观看人数达5053万,打破了此前抖音直播中的带货数据纪录。

而刘强东以数字人大佬形式的直播,也是对网友喊话“请东哥出来直播”的响应。

天眼查显示,近期,京东方面已经申请注册“老刘专场”等商标,国际分类涉及广告销售、服装鞋帽等,当前商标状态均为等待实质审查。

实际上,除了明面带货,刘强东的直播,在618大促节点前,更有卖“人”、卖“云”的意味。

在之前的预热视频中,数字人刘强东介绍自己是京东言犀AI数字人,是业内首创大佬数字人直播。

京东的官方信息显示,区别于其他的达人主播,京东采销直播以“不收坑位费、不收达人佣金”的方式。

在刘强东数字人直播后,来自京东的消息显示,目前言犀大模型已接入到京小智、虚拟主播、智能外呼等AI产品。去年双十一期间,言犀虚拟主播在超4000家品牌直播间上岗带货,并宣称应用全新技术的数字人直播确实是好生意,可复制、成本

低,可以填补真人下播后的空闲时段。

如果联想起48天前,由阿里云率先开启“击穿价”打响的那场国内云服务商的云计价价格战,当时京东云就立即回应称:“随便降,比到底。”

再联想起18天前的晚上,罗永浩卖阿里云的直播不到30分钟,就有230万人围观了。当时外界评论称,阿里云“醉翁之意不在卖货,而在占据中小客户心智”。

那么18天后,运用京东云、京东自研数字人的“采销东哥”,同样直播不到30分钟,就超过千万人围观了。

面子和里子哪个最重要?这个回答似乎并不难。



刘强东数字人直播首秀 图/视觉中国



雷军微博截图

A股春风送暖 公募积极分红

工银瑞信多措并举持续提升投资者获得感

近日,新“国九条”发布,为市场注入新动力。A股三大指数应声反弹,市场一片春意盎然。回顾2024年一季度,A股走出一波“小阳春”,银华证券数据显示,上证综指、中证全债年内分别上涨2.23%和2.34%。公募基金持续发挥专业投资服务优势,通过积极分红回馈投资者、打造长期稳健回报等方式,提升投资者获得感。

基金分红无疑是满足投资者投资需求、提升投资者获得感的重要方式。在持续波动的市场环境尤为如此。根据基金年报统计,2023年共有135家基金公司派发“红包”逾2200亿元,分红次数达6196次,其中工银瑞信旗下54只基金年度累计分红达195次,有15只基金年度分红额度超过1亿元。

作为多元发展、多点开花的基金大厂,工银瑞信建立了涵盖五条产品线的“指数家族”,全面覆盖宽基、行业主题、港股、多元配置以及指数增强等多个领域,为投资者提供了从境内到境外、从稳健到进取的全方位投资选择。银华证券数据显示,在央企改革东风再起的行情下,工银上证50ETF、工银上证50ETF联接基金均跻身同类前十,分别7/144、8/329、6/107,助力投资者精准把握主题行情。

在激烈的市场竞争中,拥有强大投研团队的基金公司,往往能够更敏锐地洞察市场动态,捕捉投资机会,并有效控制相关风险。面对近年来市场波动,工银瑞信始终坚守“投资为民”初心,以投资者利益为出发点。公司不仅积极响应监管政策,主动调低管理费和托管费,还推出了管理人合理让利型产品,为客户节省成本的同时,切实优化投资者的持有体验。(广告)